DOCUMENT RESUME

ED 322 166 TM 015 282

AUTHOR Bernstein, Lawrence
TITLE Developing an Adequately Specified Model of State Level Student Achievement with Multilevel Data.
PUB DATE Apr 90
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Educational Policy; Elementary Education; *Elementary School Students; Grade 3; Grade 5; Mathematical Models; Mathematics Achievement; *Predictive Measurement; Reading Achievement; Regression (Statistics); *School Districts; School Effectiveness; *State Programs; Testing Programs
IDENTIFIERS *Hierarchical Linear Modeling; *Multilevel Analysis; Parametric Analysis; Pennsylvania; Tests of Essential Learning Skills

ABSTRACT

Limitations of using linear, unilevel regression procedures in modeling student achievement are discussed. This study is a part of a broader study that is developing an empirically-based predictive model of variables associated with academic achievement from a multilevel perspective and examining the differences by which parameters are estimated in a multilevel analysis using regression-based hierarchical linear modeling. An introduction to hierarchical linear modeling is provided. The data for this study comprises the cohort of Pennsylvania elementary school students who were enrolled in grade 3 in 1986 and grade 5 in 1988 and who participated in the statewide Tests of Essential Learning Skills program in reading and mathematics. Results indicate the importance of correctly partitioning variability among schools or school districts; this finding has clear implications for educational policy research. Analyses using district-level aggregate data can only explain a finite portion of the variation in achievement. Although the individual-level model explains variation in student-level achievement, it cannot model policy manipulable district-level factors without biasing other effects in the model. Multilevel analysis, on the other hand, offers a means of improving conventional regression analyses confounded by aggregation bias and model misspecification. Four data tables are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LAWRENCE BERNSTEIN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED322166

Developing an Adequately Specified Model of State Level Student Achievement with Multilevel Data

Lawrence Bernstein

University of Pittsburgh

Paper presented at the annual meeting of
the American Educational Research Association
Boston, MA

April 20, 1990

TM 015282

BEST COPY AVAILABLE

I. Introduction

The report Equality of Educational Opportunity (Coleman et al., 1966) questioned the effect of schooling on student achievement and spawned numerous studies (see Glasman & Biniaminov, 1981; Madaus, Airasian & Kellaghan, 1980; and Mosteller & Moynihan, 1972 for examples) aimed at investigating the relationship between a set of input variables and student achievement within a given school system. The studies comprising the "school effects" literature have often arrived at conflicting conclusions due to the use of divergent sampling procedures, different units of analysis, disparate data analysis techniques, and varying operational definitions of the input variables and/or the outcome measures. One of the most glaring problems has been the failure to take into account the multilevel nature of the data inherent in most educational settings. For the field of educational research, this has resulted in the absence of a consensus as to what factors promote student achievement. From an educational policy perspective, this lack of consensus has led to inconsistent and counterproductive applications of empirical inquiry revolving around a number of policy questions such as school district consolidation or equity in school financing.

Recent studies (e.g., Bidwell & Kasarda, 1975; Friedkin & Necochea, 1988; Turner, Camilli, Kroc & Hoover, 1986; Walberg & Fowler, 1987) have investigated the relationship between so-called input variables, notably socioeconomic status, district size and expenditures, with student achievement at the district level within a given state. The results of these studies suggest that the role of input variables in explaining variation in student achievement is largely a function of the level of data aggregation (see Blalock, 1964; Hannan, 1971; Langbein, 1977).

Thus, proper specification of a predictive model of student achievement entails not only the inclusion of relevant predictor variables related to achievement (Cooley, 1978), but an incorporation of the multilevel nature of the data as well. In other words, developing an adequate predictive model of student achievement requires that the multilevel nature of a state educational system be taken into account by indicating how variables at one level of the system (e.g., district) might interact with variables at another level (e.g., student).

For educational research to have policy relevance the methodological mismatch between the hierarchical (i.e., nested/multilevel) nature of most educational data and the use of linear unilevel data-analytic models needs to be resolved. Previous studies of this scope and magnitude have ignored the multilevel nature of educational data, and consequently have drawn causal inferences which have obscured rather than clarified specific policy issues. Cronbach (1976) was one of the first investigators of the multilevel phenomenon to put it bluntly:

> The majority of studies of educational effects -- whether classroom experiments, or evaluations of programs, or surveys -- have collected and analyzed data in ways that conceal more than they reveal. The established methods have generated false conclusions in many studies (p. 1).

Failure to take the multilevel nature of the data into account can lead to incomplete or incorrect empirical inferences with undesirable consequences for informing state-level policy questions. In the past, the use of available aggregate-level data has been justified on the grounds that collecting individual-level information was too costly (Langbein, 1977). But as Aitkin and Longford (1986) ask:

> If the analysis of available aggregate data leads to wrong conclusions and disastrous educational policies, where is the economic or educational benefit to be found? (p. 42).

Questions of research design also have consequences regarding the issue of multilevel data. In investigating the impact of particular variables within a policy context, e.g., the effect of school district size or educational expenditures on student achievement, it is unfeasible and unrealistic to utilize experimental design methodology requiring random assignment. For example, students cannot be randomly assigned to school districts of varying size or rate of educational expenditures. Thus, educational evaluation or decision-oriented research efforts often must rely on naturally occurring data to investigate the role of a group of input variables in explaining variation in student achievement (Cooley, 1978; Longford, 1989). Unfortunately, the multilevel nature of educational data is not adequately addressed with conventional methodologies (Burstein, 1980; Cooley, Bond & Mao, 1981; Cronbach, 1976).

II. Purpose of Research

This paper is part of a broader study which has two overlapping purposes. One is to develop an empirically based predictive model which identifies the relevant variables associated with student achievement from a multilevel perspective, i.e., students within districts. The second purpose is to examine the differences in which parameters are estimated in a multilevel analysis using a regression-based hierarchical linear modeling procedure, HLM (Raudenbush & Bryk, 1986, 1988) in comparison to conventional analyses conducted at a single level of measurement (e.g., individual student or school district).

This paper addresses in part these purposes by discussing the limitations of using linear unilevel regression procedures in modeling student achievement. An introduction to the hierarchical linear modeling procedure is also provided as a brief illustration of the potential applications and benefits of multilevel analysis.

II. Data Base

The data base for this study comprises the cohort of Pennsylvania elementary school pupils who were enrolled in third grade in 1986 and in fifth grade in 1988, and who participated in the statewide Tests of Essential Learning Skills (TELS) program in reading and mathematics.[1] The analyses described in this paper are based on a total population of 86,227 students enrolled in 1804 elementary schools (in 1988) within a total of 496 school districts throughout the state of Pennsylvania. The outcome measure for this study is a composite academic achievement score (ACH88) comprised of a weighted average of the fifth grade reading and mathematics test scores. The tests are considered to be criterion-referenced measures since they indicate whether students have met state-mandated minimum competency levels in basic learning skills in reading and mathematics.

In addition to the outcome measure, several other variables are measured at the student level. A key variable is a measure of prior student ability, a composite achievement score (ACH86) based on prior performance on the third grade TELS reading and math measures. Other student background variables are SEX (1=male/2=female), RACE (1=nonwhite/2=white), special education status (SEC88), compensatory education participation in reading and math (CEPRM88), and a home environment-motivation index (HOMEMOTV).

Variables at the district level are grouped into several categories. Four district level proxies for socioeconomic status are included: AIDRATIO, a measure of a market value/personal income aid ratio for 1988; AFDCPCT, the percentage of families receiving federally-sponsored assistance for dependent children; AVGHOMTV, an aggregated mean of the home environment/motivation index; and ATTEND, a measure of average daily attendance divided by enrollment.

School district size is represented by an enrollment variable (LNADM) measuring average daily membership over the school year 1987-88. A logarithmic transformation is used to correct for nonlinearity of district size effect estimates due to the presence of a few extremely large school districts. An additional size variable (POPMILE) measures population density per square mile to reflect an urban vs. rural distinction among school districts.

Expenditures for schooling are measured by a number of 1987-1988 fiscal indicators. Instructional cost per pupil (EXPEND) is derived by dividing the net instructional expenditures of a district by average daily membership. Variables related to expenditures for schooling at

---

[1]This study draws upon a data base that has been established by Pennsylvania Educational Policy Studies (PEPS) under the direction of William Cooley. PEPS is a joint effort of the University of Pittsburgh's School of Education and the Learning Research and Development Center acting in cooperation with the Pennsylvania Department of Education.

the district level are represented by average district class size (CLSIZ), mean elementary school teacher salary in the district (TCHSAL), and ratio of the number of students to the number of district administrative personnel (ADMIN). In addition, variables measuring teacher experience (TCHSERV) and level of training (TCHLEVEL) are included.

Finally, a number of aggregated measures are included to examine the degree of contextual effects. PINSE88 measures the effect of the percentage of special education students in the district. PCTNW is the percentage of nonwhite (black and Hispanic) students in the district. DCEPRM88 is an aggregated measure of compensatory educational status at the individual student level. An additional contextual effect (DCUTTELS) measures the aggregated mean score of students falling below the cutoff on either reading or mathematics in a school district.

III. Data Analytic Procedures

The empirical development of a predictive model of student achievement was based on a series of individual student and district-level ordinary least squares regression analyses. In the district-level analysis the outcome measure (DACH88) was created as an aggregated mean of district-level achievement and regressed on a set of district-level indicators and aggregate measures of the student background variables. An individual level regression analysis was conducted with the composite achievement score (ACH88) and the student-level predictor variables.

Informed by the above results, a set of analyses were then performed using the hierarchical linear modeling program, HLM. In its simplest form, a hierarchical linear model is comprised of two equations describing the structural relationships in a within-and a between-group model (see Raudenbush and Bryk, 1986, 1988). The regression coefficients estimated in the within-group model, β_{jk}, (for k predictor variables in j groups) constitute the "microparameters" in the HLM formulation, which are then modeled as the outcome variables in the between-group model. At this next level in the hierarchy the β_{jk} regression coefficients are posited to vary across groups as a function of "macroparameters", θ_{pk}, which represent the systematic effects of **p** group-level variables on the **k** within-group relationships. Of key interest in HLM is the capturing of the variation of the structural relationships across groups through the creation of both a within- and a between-group model. HLM uses this information from both levels of the model (e.g., individual students and school districts) and therefore does not force researchers to choose between the two units of analysis.

IV. Results of District- and Individual-Level Analyses

In the district level analysis, an initial set of 16 predictors with DACH88, fifth grade achievement, as the outcome variable was entered into a multiple regression analysis for the purpose of determining the relative importance of the input variables.[2] In order to reach a more parsimonious and interpretable model, variables with insignificant regression coefficients were deleted until a model was reached with five predictors: DACH86, AVGHOMTV, AFDCPCT, LNADM, and CLSIZ (prior ability, average home environment and motivation, percent on AFDC, district size and average class size). Table 1 presents the results from the district-level analysis. The zero-order Pearson correlation coefficient with DACH88 is represented by r, while *B* and Beta refer to the unstandardized and standardized regression coefficients respectively.

Table 1

Regression Analysis for District-Level Model

Multiple R	.719
R Square	.518
Adjusted R Square	.513
Standard Error	2.536

VARIABLE	r	*B*	BETA	T	P
DACH86	.672	.487	.490	12.331	.000
AVGHOMTV	.513	1.849	.184	4.537	.000
AFDCPCT	-.458	-.083	-.175	-4.727	.000
LNADM	-.034	-.414	-.077	-2.329	.020
CLSIZ	-.102	-.386	-.109	-3.407	.001

In this model which accounted for 51 percent of the observed variance in achievement, prior ability (DACH86) not unexpectedly accounted for the largest portion of the explained variance. After prior ability is controlled for, the other four variables contributed only an additional three percent to the overall R^2. AFDCPCT displayed a moderate negative relationship with achievement, with LNADM and CLSIZ showing a weak negative relationship.[3]

---

[2]It must be emphasized here that the model-building strategy employed here is primarily of an empirical nature to allow for comparisons with a multilevel analysis. The use of mechanical procedures for variable selection in regression analysis has justifiably been criticized for its short-sightedness (see Boyd & Iversen, 1979 for examples).

[3] LNADM and CLSIZ demonstrated a "suppressor" effect in that their partial-order correlation coefficients are higher than their zero-order correlations with achievement.

AVGHOMTV exhibited a moderate positive relationship in accordance with its zero-order correlation with achievement.

What is the effect of deleting prior ability from the model? In a model without DACH86, the regression coefficients for LNADM, AVGHOMTV, CLSIZ and AFDCPCT all increased in magnitude to varying degrees. Removing prior ability from the model not only reduces the Adj. R^2 from .51 to .36, but results in affecting the magnitude of the other parameters in the model as well.

For purposes of comparison, a comparable set of analyses were conducted on the individual student level variables and achievement. A multiple regression analysis was run with ACH88 as the outcome variable and six input variables, ACH86, CEPR88, RACE, SEC88, SEX and HOMEMOTV. This model explained about 64 percent of the variance in achievement. Prior ability (ACH86) was again the strongest predictor in terms of its contribution to explained variance. SEC88 and CEPR88 had negative regression coefficients, while RACE and HOMEMOTV were positively related to achievement. There was little difference in achievement due to the effect of SEX. On the basis of these results, SEX was dropped from the analysis and the individual level model was refitted with the remaining five predictors. This final model did not suffer any drop in the adjusted R^2 figure and the model estimates remained virtually identical. Table 2 presents the results for this model with correlation coefficients (r) and unstandardized (B) and standardized (Beta) regression weights.

Table 2

Regression Analysis for Individual-Level Model

Multiple R	.797
R Square	.636
Adjusted R Square	.636
Standard Error	8.987

VARIABLE	r	B	BETA	T	P
ACH86	.776	.675	.622	230.687	.000
HOMEMOTV	.349	.875	.082	37.224	.000
RACE	.342	5.037	.109	49.883	.000
CEPRM88	-.460	-3.462	-.122	-52.014	.000
SEC88	-.325	-6.854	-.095	-42.774	.000

As with the district level regression, an analysis was run without ACH86 to investigate the effect of deleting prior ability on model specification. An analysis with the remaining four

predictor variables had an adjusted R^2 figure of .41, a decrease of about one-third from the previous model. In this model, all the effects increased dramatically in size and remained in the same direction.

Analyses conducted on the individual level can account for variables measured at a higher level of aggregation by disaggregating the effects back to the individual level (see Summers & Wolfe, 1977 for an example). Contextual effects can thus be estimated by including a constant value associated with the school district for each individual student record. A model including the contextual effects of percent nonwhite (PCTNW), average home environment/motivation (AVGHOMTV), percent in compensatory education (DCEPRM88), below cutoff in reading/math (DCUTTELS), and percent in special education (PINSE88) increased the explained variability in achievement by only two percentage points to about 65 percent. Of more importance, however, is the interpretation of the regression coefficients in this model.

In this new model with district level contextual effects predicting student achievement, the individual level effects remained the same, except for race, which decreased by more than 50 percent. The coefficients for the contextual effects, PCTNW, PINSE88 and DCEPRM88 were all positive whereas their zero-order correlations with achievement were negative. AVGHOMTV, on the other hand, had a negative effect on achievement in this model as opposed to its positive zero-order correlation with ACH88. Only DCUTTELS had the expected effect: The higher the district mean score of students below the cutoff on reading and math, the lower an individual student's score.

The explanation for this counter-intuitive pattern of results lies in the simultaneous modeling of a group of individual and group-level effects. Not only are the individual variables correlated with their district-level counterparts, but the district-level variables are also intercorrelated with each other.[4] Multicollinearity within the context of simultaneously modeling individual and group-level effects creates a difficulty in interpreting the relative importance of the explanatory variables in the model (Boyd and Iversen, 1979).

The resolution of this problem lies in adopting a technique which allows for the explicit modeling of individual-level relationships as a function of group-level factors. The following section of the results is devoted to a review of the analyses employed with the hierarchical linear modeling (HLM) program.

---

[4]The correlations among the district-level contextual effect variables at the individual level are higher than their corresponding intercorrelations at the district level.

V. Results of Multilevel Analysis

An HLM model was formulated for the purposes of analyzing district mean variability. In this analysis, information was initially provided as to how much variation in the outcome measure, ACH88, lay between and within districts. The model, equivalent to a one-way analysis of variance with districts treated as a random factor was posed as:

$Y_{ij} = \mu_j + \varepsilon_{ij}$, (within-district) and

$\mu_j = \theta + U_j$ (between-district)

In this formulation, the within-district model states that the outcome variable Y_{ij} (fifth grade academic achievement of student i in district j) varies around a district mean μ_j with independent errors ε_{ij} distributed $N(0,\sigma^2)$ where σ^2 signifies within-district variance. In turn, in the between-district model, each district's mean μ_j varies around a grand mean θ with independent errors U_j distributed $N(0,\tau)$ with τ signifying between-district variance.

The ratio of τ (between-district variance) to $\tau + \sigma^2$ (between- and within-district variance) yields an intra-district correlation coefficient $\hat{\rho}$, which expresses the estimated proportion of variance in the outcome measure between districts. In this study, $\tau = 11.28$ and $\sigma^2 = 185.09$, with $\hat{\rho} = .057$ indicating that approximately only six percent of the variance in Y is located between districts.

The next question concerned whether district means varied significantly across districts, or H_0: $\tau = 0$, where τ again represents the amount of between-district variation. In a test of this hypothesis, the Chi-square test statistic was equal to 19132.0 with 495 degrees of freedom, $p<.001$, indicating that the null hypothesis could be rejected and that districts did show significant variability in mean achievement.

The final question centered around determining the contribution of district-level factors to explaining the variability in district mean achievement. In this between-district model, each district mean score is predicted by district factors such as district mean home environment/motivation (SES):

$\mu_j = \theta_0 + \theta_1(\text{mean SES})_j + U_{0j}$,

where

θ_0 is equal to the grand mean of achievement,

θ_1 is equal to the effect of district mean SES on μ_j,

and

U_{0j} is distributed $N(0,\tau)$, where τ now represents the residual parameter variance after controlling for district mean SES.

The results showed a highly significant relationship between district SES as measured by home environment-motivation and district mean achievement ($\theta_1 = 5.12$, $t = 13.26$, $p<.001$).

The residual parameter variance, τ, after accounting for the district SES factor is now reduced to 7.90 from 11.28, a reduction of 30 percent. In other words, about 30 percent of the between-district variance in mean achievement is accounted for by district home environment/motivation. In this model, however, mean achievement still varied significantly across districts (Chisquare = 13013 with 494 df, $p<.001$). This indicates that more terms need to be added to the between-district model in order to account for the variation in district mean achievement.

Table 3 displays the results from an analysis using the variables from the district-level model with θ coefficient values, standard errors and t-ratios:

Table 3

HLM Results for Model with Five District-Level Variables

Residual Parameter Variance			4.33	
R Square			.62	
VARIABLE	**θ**	**S.E.**	**T**	**P**
DACH86	.492	.039	12.730	.000
AVGHOMTV	1.726	.396	4.361	.000
AFDCPCT	-.085	.017	-4.995	.000
LNADM	-.436	.173	-2.516	.012
CLSIZ	-.361	.111	-3.266	.001

In this model after accounting for the effects of prior ability, average home environment-motivation, percent on AFDC, district size and average class size, the residual parameter variance is reduced to 4.33, a reduction of 62 percent over the unconditional model with no district effects included. The θ effect for prior ability (DACH86) indicates a highly significant association with district mean achievement. The other four predictors, to a lesser extent, are also significantly related to explaining variation in district mean achievement as indicated by their θ coefficients. In this model, however, residual parameter variance in mean achievement still varied significantly across districts (Chisquare = 3352.8, df=490, $p<.001$).

The strong effect of prior ability on district mean achievement mediates the effects of the other district-level predictors.[5] In an attempt to assess the potential effects of these

[5]In fact, a between-district model containing only DACH86 resulted in a reduction of 54 percent in terms of explaining parameter variance.

variables, a model was formulated without the district-level effect of prior ability DACH86. The results of this model are shown in Table 4:

Table 4

HLM Results for Model without Prior Ability

Residual Parameter Variance			6.31	
R Square			.44	
VARIABLE	**θ**	**S.E.**	**T**	**P**
AVGHOMTV	4.051	.406	9.983	.000
AFDCPCT	-.145	.019	-7.643	.000
LNADM	-.573	.200	-2.865	.005
CLSIZ	-.530	.127	-4.170	.000

In this new model, the θ coefficients for the four district-level variables increase considerably. The estimated residual parameter variance, τ, for this model is 6.31 which represents a reduction of 44 percent over the unconditional model. Mean achievement, however, still varies significantly across districts (Chisquare = 6302.7 with 491 df, p<.000). This test statistic indicates that more variables need to be added to the model to explain the remaining variation in achievement.

The question remains as to how an analysis employing multilevel data differs from a conventional analysis conducted at, for example, the district level. These results from Tables 1 and 3 show a striking similarity between the estimates obtained from the ordinary least squares regression district-level analysis and from the hierarchical linear modeling analysis. This is not at all surprising given the fact that in both instances mean district achievement is being predicted by the same set of input variables. The multilevel analysis, however, accounted for a greater share of the explained variance in achievement (62 percent) than the OLS analysis at the district level (51 percent). This is explained by the fact that the hierarchical linear modeling procedure partitions the variance in district means into parameter variance and sampling variance as opposed to the conventional regression analysis which does not make this distinction. Thus, in the HLM case, the set of five predictor variables explained 62 percent of the "true" differences in district achievement potential of explanation and not attributable to sampling variability.

The story at this stage is by no means complete. Further analysis will focus on the study of the variability in within-district slopes modeled as a function of district-level

characteristics. For example, the relationship between within-district SES and achievement will be examined as a function of district-level variables such as district size. One of the questions to be examined here is whether large school districts tend to have a more pronounced SES effect on student achievement than small ones. This "interaction" effect would indicate perhaps that low SES students are disadvantaged even further in large school districts.

VI. Summary

The above results suggest various implications for policy research, particularly in terms of how policy can get mislead by inappropriate methodological models given the hierarchical nature of educational data. For one, the importance of correctly partitioning variability among schools or school districts has clear implications for educational policy research. As indicated in the results only a small portion of variability in student achievement is potentially explainable by group-level factors. Analyses employing district-level aggregate data, thus, are only capable of explaining a finite portion of the variation in achievement. District-level analyses explain **district-level** achievement.[6] The district-level model in this study with an R^2 of .51 is only explaining, in fact, 51 percent of the potentially explainable portion (in this case six percent) of the variation in student achievement.

The individual level model, on the other hand, had a higher R^2 of .64 primarily due to the stronger relationship between third and fifth grade achievement at the student level. Individual level models, however, while explaining variation in **student-level** achievement, are incapable of modeling policy manipulable district-level factors without biasing other effects in the model. The purpose of the above comparisons was not to demonstrate the superiority of one level of analysis over the other. The question of which level of analysis to choose is clearly the wrong question to ask. In this case, the lack of a strong grouping effect at the school district level would lead one instead to look at potential classroom- or school-level factors to explain variability in student achievement.

These analyses have demonstrated, however, whether one is conducting a multilevel analysis or a conventional ordinary least squares regression at the district or individual level, the importance of specifying prior ability in a model to reduce the bias due to student self-

---

[6]Of that portion, it is crucial to identify those manipulable policy variables which have a potential impact on individual student achievement (Langbein, 1977).

selection.[7] Analyses employing proxies such as socioeconomic status for incoming student ability are simply inadequate for addressing this bias. Better and more complete estimates of student attributes mitigating the problems of missing data and non-response, as well as the employment of longitudinal data to measure long-term educational effects, are other concrete examples of developing more adequately specified models of student achievement.

It is not expected that this approach using hierarchical linear modeling can provide a conclusive conceptual answer to the question of correct model specification of student achievement. Due to the non-random manner in which students are grouped into schools and districts, model misspecification in terms of biased parameter estimates will continue to be a problem, no matter how sophisticated the statistical methodology employed. By specifying the manner in which variables measured at different levels are related, however, multilevel analysis offers hope of an improvement over conventional regression analyses confounded by aggregation bias and model misspecification.

[7]Conditioning on prior ability **reduces** but cannot **eliminate** initial differences in academic achievement.

References

Aitkin, A., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society,* (Series A), *149,* 1-43.

Bidwell, C.E. & Kasarda, J.D. (1975). School district organization and student achievement. *American Sociological Review, 40,* 55-70.

Blalock, H.M. (1964). *Causal Inferences in Nonexperimental Research.* Chapel Hill: University of North Carolina Press.

Boyd, L.H. & Iversen, G.R. (1979). *Contextual Analysis: Concepts and Statistical Techniques.* Belmont, CA: Wadsworth.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education, 8,* 158-233.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D. & York, R.L. (1966). *Equality of educational opportunity.* Washington, D.C.: Office of Education, U.S. Department of Health, Education and Welfare.

Cooley, W.W. (1978). Explanatory observational studies. *Educational Researcher, 7,* 9-15.

Cooley, W.W., Bond, L., & Mao, B-J. (1981). Analyzing multilevel data. In R.A. Berk (Ed.), *Educational evaluation methodology: The state of the art.* (pp. 64-83). Baltimore: The Johns Hopkins University Press.

Cronbach, L.J. (1976). *Research on classrooms and schools: Formulation of questions, design and analysis.* Occasional paper of the Stanford Evaluation Consortium.

Friedkin, N.E. & Necochea, J. (1988). School system size and performance: A contingency perspective. *Educational Evaluation and Policy Analysis, 10,* 237-249.

Glasman, N.S. & Biniaminov, I. (1981). Input-output analyses of schools. *Review of Educational Research, 51,* 509-539.

Hannan, M.T. (1971). Problems of aggregation. In H.M. Blalock (Ed.), *Causal models in the social sciences.* (pp. 473-508). Chicago: Aldine-Atherton, Inc.

Langbein, L.I. (1977). Schools or students: aggregation problems in the study of student achievement. In M. Guttentag (Ed.), *Evaluation Studies Review Annual, Vol. 2.* (pp. 270-298). Beverly Hills, CA: Sage.

Longford, N.T. (1989). Variance and component analysis: Technical appendix. In D.J. Smith & S. Tomlinson (Eds.), *The school effect.* London: Policy Studies Institute.

Madaus, G.F., Airasian, P.W., & Kellaghan, T. (1980). *School effectiveness: A reassessment of the evidence.* New York: McGraw-Hill.

Mosteller, F. & Moynihan, D.P. (1972). *On equality of educational opportunity.* New York: Random House.

Raudenbush, S.W. & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59,* 1-17.

Raudenbush, S.W. & Bryk, A.S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E.Z. Rothkopf (Ed.), *Review of Research in Education, Vol. 15.* (pp. 423-475). Washington, D.C.: AERA.

Summers, A.A. & Wolfe, B.L. (1977). Do schools make a difference? *The American Economic Review, 67,* 639-652.

Turner, R., Camilli, G., Kroc, R., & Hoover, J. (1986). Policy strategies, teacher salary incentive, and student achievement: An explanatory model. *Educational Researcher, 15(3),* 5-11.

Walberg, H.J. & Fowler, W.J. (1987). Expenditure and size efficiencies of public school districts. *Educational Researcher, 16,* 5-13.